

Validation of the World Health Organization Tool for Situational Analysis to Assess Emergency and Essential Surgical Care at District Hospitals in Ghana

Hayley Osen · David Chang · Shelly Choo ·
Henry Perry · Afua Hesse · Francis Abantanga ·
Colin McCord · Kristin Chrouser · Fizan Abdullah

Published online: 29 December 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract

Background The World Health Organization (WHO) Tool for Situational Analysis to Assess Emergency and Essential Surgical Care (hereafter called the WHO Tool) has been used in more than 25 countries and is the largest effort to assess surgical care in the world. However, it has not yet been independently validated. Test–retest reliability is one way to validate the degree to which tests instruments are free from random error. The aim of the present field

study was to determine the test–retest reliability of the WHO Tool.

Methods The WHO Tool was mailed to 10 district hospitals in Ghana. Written instructions were provided along with a letter from the Ghana Health Services requesting the hospital administrator to complete the survey tool. After ensuring delivery and completion of the forms, the study team readministered the WHO Tool at the time of an on-site visit less than 1 month later. The results of the two tests were compared to calculate kappa statistics for each of the 152 questions in the WHO Tool. The kappa statistic is a statistical measure of the degree of agreement above what would be expected based on chance alone.

Results Ten hospitals were surveyed twice over a short interval (i.e., less than 1 month). Weighted and unweighted kappa statistics were calculated for 152 questions. The median unweighted kappa for the entire survey was 0.43 (interquartile range 0–0.84). The infrastructure section (24 questions) had a median kappa of 0.81; the human resources section (13 questions) had a median kappa of 0.77; the surgical procedures section (67 questions) had a median kappa of 0.00; and the emergency surgical equipment section (48 questions) had a median kappa of 0.81.

Conclusions Hospital capacity survey questions related to infrastructure characteristics had high reliability. However, questions related to process of care had poor reliability and may benefit from supplemental data gathered by direct observation. Limitations to the study include the small sample size: 10 district hospitals in a single country. Consistent and high correlations calculated from the field testing within the present analysis suggest that the WHO Tool for Situational Analysis is a reliable tool where it measures structure and setting, but it should be revised for measuring process of care.

Electronic supplementary material The online version of this article (doi:10.1007/s00268-010-0918-1) contains supplementary material, which is available to authorized users.

H. Osen (✉) · D. Chang
Department of Surgery, Center for Surgical Systems and Public Health, University of California San Diego, San Diego, CA, USA
e-mail: HBOsen@gmail.com

S. Choo · F. Abdullah
Department of Surgery, Johns Hopkins University, Baltimore, MD, USA

H. Perry · K. Chrouser
Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

A. Hesse
Department of Surgery, Korle Bu Teaching Hospital, Accra, Ghana

F. Abantanga
Department of Surgery, Komfo Anokye Teaching Hospital, Kumasi, Ghana

C. McCord
Department of Surgery, Columbia University, New York, NY, USA

Introduction

In a 2008 article, Farmer and Kim [1] aptly described surgery as the “neglected stepchild” of global public health. Indeed, an estimated 234 million surgical procedures occur each year [2], and surgical disease accounts for 15% of the total disability adjusted life-years (DALYs) lost worldwide [3]. Despite surgical care being demonstrated to be highly cost effective [3], surgical care in low- and middle-income countries (LMICs) continues to be underfunded as a public health priority.

Although many studies have assessed the surgical need for a specific area or population, the findings of these studies may be difficult to generalize and have limited usefulness for policy makers. The World Health Organization (WHO) Tool for Situational Analysis to Assess Emergency and Essential Surgical Care (hereafter referred to as the WHO Tool (<http://www.who.int/surgery/publications/QuickSitAnalysisEESCSurvey.pdf>) was developed in 2007 as part of the research group of the WHO Global Initiative for Emergency and Essential Surgical Care (GIEESC) [4]. The data are entered into the WHO database and are shared with health authorities for evidence-based decisions on planning. The objective of using this tool is to provide a cross-sectional assessment of the state of surgical care in a given hospital, identifying the strengths, weaknesses, and gaps in four key aspects of the surgical health care delivery system: infrastructure, human resources available to provide surgical care, surgical interventions being performed, and emergency equipment available for the care of surgical patients. Since its inception, the Tool for Situational Analysis has been used in more than 25 countries, making it the most widely used questionnaire to assess surgical capacity in the world.

Despite its widespread use, the WHO Tool has never been independently validated. In particular, no statistical tests have been conducted to show that the WHO Tool actually measures what it was intended to measure. Test–retest reliability is one way to validate the degree to which test instruments are free from random error. The aim of the present field study was to determine the test–retest reliability of the WHO Tool.

Methods

The study population included a convenience sample of 10 district hospitals in Ghana, with one hospital selected for each region in Ghana: Hohoe Hospital (Volta Region), Winneba Hospital (Central Region), Goaso Hospital (Brong Ahafo Region), Tumu Hospital (Upper West Region), Sadema Hospital (Upper East Region), Begoro Hospital (Eastern Region), Bekwai Hospital (Ashanti

Region), Dodowa Hospital (Greater Accra Region), Bibiani Hospital (Western Region), and Bole Hospital (Northern Region). District hospitals in developing countries play an important role as the first level of referral for patients presenting with surgical and obstetric conditions, such as obstructed labor and acute surgical abdomen. Altogether, there are 124 district hospitals in Ghana. Each region also has a regional hospital, which has better surgical capability but can be difficult to access for those living in rural and remote areas.

The study team included Ghana health authorities and representatives of international and local academia. The questionnaire was administered by mail to the hospital administrators, who were asked to complete it prior to study team arrival. The in-country study team member called each of the hospitals to confirm receipt and completion of the survey tool prior to study team visit. Upon arrival, the study team readministered the same questionnaire a second time in person to the same hospital administrator who had completed the survey prior to the study teams onsite visit.

The results of the two administrations of the questionnaires were then compared, and kappa statistics were calculated. The kappa statistic is a measure of the degree of agreement above the expected agreement based on chance alone for categoric responses. It can be expressed in the following formula [5] and further explained by Fig. 1

$$\kappa = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}}$$

Kappa values > 0.8 are considered almost perfect; at 0.6–0.8, substantial agreement; at 0.4–0.6, moderate agreement; at 0.2–0.4, fair agreement; and < 0.2, slight agreement [5]. As a sensitivity analysis, the weighted kappa was calculated. This value accounted for there being more response categories for some sections (e.g., processes section) and fewer response categories for other sections (e.g., infrastructure section).

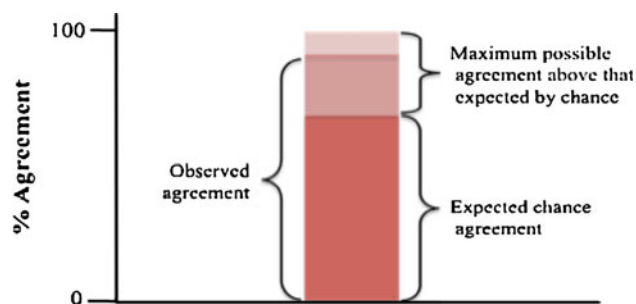


Fig. 1 Middle portion of the bar indicates the quantity captured by the kappa statistic: the observed agreement above that expected by chance

The kappa statistic was not calculated for questions that may not have been answered by all respondents. For example, in the “Procedures” section, there were two “required” questions along with three conditional questions: Respondents would have answered the last three columns (“Do you refer due to lack of skills,” “...due to nonfunctional equipment,” “...due to lack of supplies”) only if they had answered “yes” to “Do you refer?” Therefore, Table 1 shows that the “Procedures” section had 70 questions (35×2) rather than 175 (35×5). For some questions, kappa could not be calculated. kappa cannot be calculated when there is unanimous agreement among all respondents. This is because the expected probability is calculated by the likelihood of a certain response on first administration, multiplied by the likelihood of that response on second administration. Therefore, if the probability of selecting answer A is 100% on first administration (i.e., every respondent selected answer A) and also 100% on second administration (i.e., every respondent selected answer A again), the expected probability is 100%, which means there is no room for any unexpected agreement; hence, kappa cannot be calculated. In survey methodology terminology, these questions essentially have a floor or a ceiling effect and they are nondiscriminatory and thus provide no informational value concerning test–retest reliability.

Results

All 10 surveys were successfully completed twice (100% completion rate): once before arrival of the study team, and a second time after the arrival of the onsite study team at the hospital. kappa could be calculated for 81.7% of the 186 questions. As noted above, kappa could not be calculated for 34 questions because of the unanimity of responses for those questions. The overall median unweighted kappa for all of the 152 questions analyzed was 0.43 (interquartile range 0–0.84), and the overall median weighted kappa was 0.43 (0–0.89), indicating that there

was no significant difference between the unweighted and weighted values. The number of kappas calculated for each section and the median unweighted and weighted kappa values for each section are shown in Table 1.

Discussion

Major health care issues can be broadly categorized into availability, access, utilization, quality, and outcomes [6]. Research on surgical capacity in LMICs is lacking [7], but studies that have been done focused on availability and access issues, rather than quality of care, which seems to have been reserved as an achievable metric in better-resourced health systems. However, Ghana has established a relatively good health care system because of its investments in primary health care, hospital referral systems, lower-level trained staff, and financial support for the system, including government-funded public health insurance [8]. Ghana is thus a good test case to begin to expand beyond issues of availability and access and start considering issues related to quality of care for key service delivery issues such as surgical care at district hospitals.

The quality of health systems can be analyzed using the Donabedian model, which systematically analyzes the health system according to three domains: structure, process, outcomes [9]. Structure refers to the physical aspects of the health delivery system, including buildings, equipment, and human capital; and this aspect of quality is often particularly important to consider in areas where infrastructure is poor. The process component of the model refers to how each component of the health care delivery system operates, including how the hospital triages patients, a patient’s hospital course, and the processes needed to make the health system function smoothly. The final “outcomes” component of the model ultimately allows one to assess the quality of the patient care produced by the health system. Whereas most studies of health systems in low-income countries focus only on structural issues (i.e., access and availability of care), the Donabedian

Table 1 kappa statistics calculation, by section of the WHO survey instrument

Section	No. of questions	No. of kappas calculated	Unweighted median kappa (IQR)	Weighted median kappa (IQR)
Infrastructure	24	24	0.81 (0.61–1.0)	0.84 (0.58–1.0)
Human resources	24	13	0.77 (0.43–1.0)	0.78 (0.17–1.0)
Procedures	70	67	0 (0–0.41)	0 (0.17–1.0)
Emergency equipment	67	48	0.81 (0–1.0)	0.84 (0–1.0)
Total	185	152	0.43 (0–0.84)	0.43 (0–0.89)

WHO World Health Organization, IQR interquartile range

model advocates for a more global approach by considering upstream issues that could affect the outcomes of care, including structure and process.

The results of this study are striking because they highlight the need to revise the WHO Tool so it better measures issues related to the processes of a system, such as the interventions provided in a hospital. In this study, we found that questions related to structural issues (infrastructure, human resources, emergency equipment) have high reliability, reflecting the clarity and usefulness of those questions. In contrast, questions related to process of care (surgical procedures performed and reasons for referral of surgical patients) have poor reliability. For example, the first question of the “procedures” section asks the administrator if the hospital provides resuscitation. The poor reliability indicates that hospital administrator may not know if the hospital provides the procedure, may think it can be provided but has not had a patient who needed it, or may have performed the procedure but did not technically have the proper personnel or equipment. Based on these data, we recommend that the procedural section of the instrument be revised, with a new set of questions that are developed with survey reliability in mind. It may be argued that this section of the survey suffers because it is overly complex and tries to capture too much data; for example, each question in this section has five subquestions. The wording of the questions may also need to be clarified; for example, the respondents told us that they were confused about the difference between “refer patients due to lack of supplies/drug” versus “refer patients due to nonfunctional equipment.” There may be a role for local focus groups in the revision process.

The strength of our study is the quantitative assessment of the test–retest reliability of the survey instrument. Our study represents the first reported validation of this important tool, which is being used in the world’s largest assessment of availability and quality of surgical services in LMICs. Additional assessment of the tool is required for its further improvement and usefulness in measuring the quality of surgical care in low-resource settings.

One of the strengths of our study was the inclusion of every region in Ghana, which increases the validity of our findings. A limitation of our study includes the fact that surgical outcomes at the hospital level could not be assessed. The study was also based on a convenience sample of hospitals, which may not be as generalizable as a randomly selected sample of hospitals. Additionally, this study focused on intrarater reliability: we surveyed the same respondents twice. It would have been valuable to assess interrater reliability as well by surveying two individuals from the same hospital. Furthermore, although it is unlikely, it is possible that some disagreement between kappa statistics occurred due to legitimate changes in

hospital structure and processes during the 1-month study period.

Future studies should examine additional approaches to assess the process of surgical care in health systems in low-resource settings. They might include direct observational studies, which may provide more reliable data on processes of care. Additionally, it may be valuable to compare results from a direct observational study versus a survey-based study to determine the validity of the survey responses.

Conclusions

Most of the sections of the WHO Tool for Situational Analysis to Assess Emergency and Essential Surgical Care are reliable for assessing the capacity of district hospitals in Ghana. The WHO Tool appears to have relatively good test–retest reliability for the sections on physical infrastructure, emergency surgical equipment, and human resources. However, the section that addresses surgical procedures performed and the frequency of referral depending on the nature of the surgical problem should be modified to improve its reliability for addressing gaps in the availability of procedures. This will be valuable to both health care providers as well as policymakers. This WHO Tool allows us to begin to examine in greater depth the capacity of a district hospital in a low-resource setting to provide emergency and essential surgical care. In addition to issues of access and availability, it is important to begin to bring quality of care issues into our assessment of surgical systems in low resource settings. Given its leadership in health care system infrastructure investment, Ghana is uniquely poised to take the lead in this future research agenda for sub-Saharan Africa.

Acknowledgments Bloomberg Philanthropies and the World Lung Foundation supported this work.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Farmer PE, Kim JY (2008) Surgery and global health: a view from beyond the OR. *World J Surg* 32:533–536
2. Weiser TG, Regenbogen SE, Thompson KD et al (2008) An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 372:139–144
3. Debas HT, Gosselin R, McCord C et al (2006) Surgery. In: Jamison D (ed) *Disease control priorities in developing countries*, 2nd edn. Oxford University Press, New York, pp 1245–1260

4. Emergency and Essential Surgical Care (2009) <http://www.who.int/surgery/globalinitiative/en/>. Accessed 5 April 2010
5. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
6. Shengelia B, Tandon A, Adams OB et al (2005) Access, utilization, quality, and effective coverage: an integrated conceptual framework and measurement strategy. *Soc Sci Med* 61:97–109
7. Bickler SW, Spiegel DA (2008) Global surgery: defining a research agenda. *Lancet* 372:90–92
8. National Health Insurance Act of 2003. <http://www.nhis.gov.gh/?CategoryID=158&ArticleID=80>. Accessed 5 April 2010
9. Donabedian A (1966) Evaluating the quality of medical care. *Milbank Mem Fund Q* 44:166–206